APPLICATION OF SPECTRAL ANALYSIS TO STREAM AND ESTUARY FIELD SURVEYS

I. Individual Power Spectra

T. A. Wastler
Technical Services Branch
Robert A. Taft Sanitary Engineering Center

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Division of Water Supply and Pollution Control

Cincinnati, Ohio
November 1963

The ENVIRONMENTAL HEALTH SERIES of reports was established to report the results of scientific and engineering studies of man's environment: The community, whether urban, suburban, or rural, where he lives, works, and plays; the air, water, and earth he uses and re-uses; and the wastes he produces and must dispose of in a way that preserves these natural resources. This SERIES of reports provides for professional users a central source of information on the intramural research activities of Divisions and Centers within the Public Health Service, and on their cooperative activities with State and local agencies, research institutions, and industrial organizations. The general subject area of each report is indicated by the two letters that appear in the publication number; the indicators are

WP - Water Supply and Pollution Control
AP - Air Pollution
AH - Arctic Health
EE - Environmental Engineering
FP - Food Protection
OH - Occupational Health
RH - Radiological Health

Triplicate tear-out abstract cards are provided with reports in the SERIES to facilitate information retrieval. Space is provided on the cards for the user's accession number and additional key words.

Reports in the SERIES will be distributed to requesters, as supplies permit. Requests should be directed to the Division identified on the title page or to the Publications Office, Robert A. Taft Sanitary Engineering Center, Cincinnati, Ohio 45226.

Public Health Service Publication No. 999-WP-7

CONTENTS

ABSTRACT

The application of spectral analysis techniques to sanitary engineering stream and estuary studies is discussed from a practical operational viewpoint. Techniques of interpretation and the data requirements are emphasized rather than the mathematical basis and details of the technique. The usefulness of spectral analysis in analyzing records obtained from continuous, automatic monitoring stations is pointed out. Spectral analyses applied to tidal height records and dissolved oxygen records obtained in a field study of the Potomac Estuary are discussed. The discussion is limited to the application of individual power spectra computation to sanitary engineering investigations.

APPLICATION OF SPECTRAL ANALYSIS TO STREAM AND ESTUARY FIELD SURVEYS

I. Individual Power Spectra

INTRODUCTION

The expanding population in the coastal areas of this continent has stimulated interest in the sanitary engineering study of estuarine systems. The speed with which conditions change in estuaries and the complex interactions that occur require new techniques of estuarine sampling and data analysis.

Spectral analysis is a data analysis tool that shows considerable promise for both estuary and stream study. This technique emerged from the work of Fourier and Laplace near the beginning of the nineteenth century, and much of the mathematical theory underlying its practical application was developed in the work of G. I. Taylor, Norbert Wiener, and S. O. Rice. It was not until the advent of automatic sampling devices and high-speed digital computers that its practical application to many scientific and engineering problems became feasible. Since the 1940's spectral analysis has been used with considerable success in such diverse fields as communications engineering, aerodynamics, oceanography, and meteorology.

The purpose of this discussion is to present the computation and interpretation of individual power spectra, the basic operation of spectral analysis, as an engineering tool. The mathematical and statistical basis of spectral analysis, which is essentially a mathematical and statistical technique, has been almost entirely ignored. It is hoped that the references given in the bibliography may satisfy any need for fuller explanation of these aspects of spectral analysis.

THE CONCEPT OF SPECTRAL ANALYSIS

One of the classic experiments in the study of optics is to pass a beam of sunlight through a triangular prism and observe the rainbow of colors into which the beam of white light is split. This experiment serves as a useful physical analogy to the operation of spectral analysis on a body of data.

In Figure 1a is presented a schematic diagram of the results of passing a beam of sunlight through a triangular prism. The beam of sunlight is split into a spectrum of colors ordered according to their respective wave lengths (or frequencies). If this spectrum of colors is allowed to strike a battery of light-sensitive cells, the intensity of the light of each color can be measured; these results can then be plotted as a "light intensity spectrum," which might look like that in Figure 1a if each of the six principal colors had the same intensity.

If a similar experiment is performed with a light beam that is composed of only three of these colors (present at different intensities), the resulting light frequency spectrum and light intensity spectrum might resemble those in Figure 1b.

This experiment demonstrates the resolution of a complex physical phenomenon into a group of simpler phenomena that may be easier to examine from both theoretical and practical viewpoints.

The effect of using spectral analysis on a record of observed field data is directly analogous to the effect of the prism on the light beam. This analogous effect is presented schematically in Figure 1c. The actual technique of spectral analysis is discussed later; at this time the significance of the result of the computation is of concern.

In the prism experiment the relative intensities of the resolved light frequencies can be observed and studied. Spectral analysis of a record of observations results in a sorting of the total "variance" of the record into its component frequencies. The variance of a data record is therfore analogous to the intensity of the light beam.

The variance is defined as the sum of the squares of the deviations from the mean divided by one less than the number of observations. This is the definition of variance as it is normally used as a descriptive statistic of a body of data. Conceptually, the variance is a measure of the dispersion of observations about the mean value. In the statistical interpretation of data, it is ordinarily regarded that this dispersion of values about the mean is due to chance.

L ANALYSIS

tudy of optics is to
- prism and observe
f white light is split.
l analogy to the oper-

diagram of the results
ngular prism. The
colors ordered ac-
frequencies). If
a battery of light-
each color can be
as a "light inten-
n Figure 1a if each
nsity.

ith a light beam that
resent at different
pectrum and light in-
igure 1b.

lution of a complex
henomena that may
and practical view-

a record of observed
of the prism on the
ted schematically in
l analysis is dis-
the result of the

ensities of the re-
l studied. Spectral
in a sorting of the
nent frequencies.
alogous to the inten-

e squares of the de-
han the number of
nce as it is normally
ata. Conceptually,
of observations about
tion of data, it is
lues about the mean

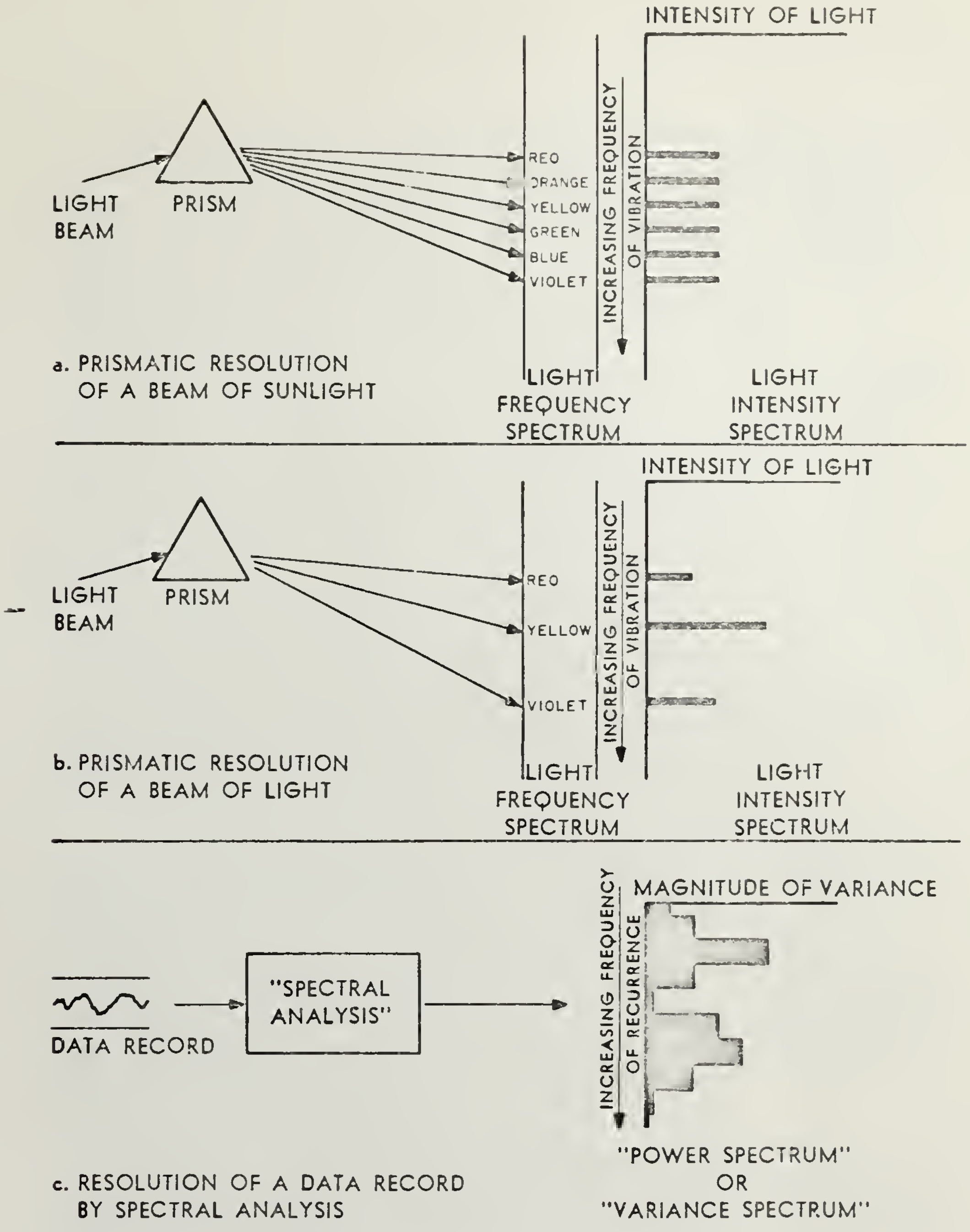


Figure 1. Physical analogy to spectral analysis.

The statistical manipulation of time-series data by spectral analysis results in the computation of those parts of the variance of a record that recur at constant time intervals as well as the part that is random (nonrecurring) in character. As the light beam may be resolved by the prism into its component colors of different intensity, so is the variance of a time-series record resolved into its component parts by spectral analysis.

The estimates of variance for each frequency resolved in the spectral analysis form the "power spectrum" of the record from which the computations are made. (The term "variance spectrum" would be more accurate; but in the pioneering work in spectral techniques done in communications engineering the term "power," which is closely related to record variance in that frame of reference, became common usage.) The power spectrum computation is the fundamental operation of data reduction and interpretation by means of spectral analysis; the many other computations that can be made in spectral analysis are all based firmly upon the concept and calculation of individual power spectra. It may be stated that the computation of individual power spectra bears about the same relationship to spectral analysis as differentiation and integration bear to the Calculus.

The interpretation of variance as a statistic descriptive of both the random and nonrandom characteristics of a time-series data record is most important in understanding the significance of spectral results. In the usual type of statistical analysis, variance is conceptually regarded as a measure of the random dispersion of the observations from their mean value. In many cases this is true; but that this is not a necessary condition for the existence of a variance can be demonstrated with the aid of Figure 2, in which segments of three hypothetical records and the corresponding power spectra are presented.

Figure 2a shows a record that has a constant value, i. e., all values are equal to the mean. Since there are no deviations from the mean, the variance is zero and the power spectrum is zero at all frequencies.

Figure 2b shows a record that forms a sloping straight line. The segment of the record shown in this figure has a mean of 3.55 and a variance of 0.69. It is apparent that none of this variance results from a "random" dispersion about the mean, but is the result of a secular (time-dependent) trend in the record. If spectral analysis were done on the record of which this segment is a part, all of the variance (or "power") would be concentrated in the zero-frequency spectral estimate as shown in Figure 2b. The zero-frequency spectral estimate includes all of the record variance that does not recur during the length of the record used in the analysis. It therefore includes (1) any truly random fluctuations in the record, (2) any linear trends in the

eries data by spectral
parts of the variance
rvals as well as the
racter. As the light
component colors of
a time-series record
tral analysis.

equency resolved in the
m" of the record from
rm "variance spec-
pioneering work in
ns engineering the term
rd variance in that
e.) The power spec-
ation of data reduction
alysis; the many other
analysis are all based
individual power spec-
n of individual power
to spectral analysis as
Calculus.

atistic descriptive of
istics of a time-series
nding the significance
t tical analysis,
asure of the random
mean value. In many
cessary condition for
trated with the aid of
thetical records and
ented.

constant value, i.e.,
ere are no deviations
he power spectrum is

a sloping straight line.
ure has a mean of
t that none of this
on about the mean,
nt) trend in the re-
e record of which
(or "power") would be
l estimate as shown in
stimate includes all of
ring the length of the
ncludes (1) any truly
linear trends in the

record, and (3) any periodic components in the record that are of so low a frequency that they appear as linear trends in the record. For example, spectral analysis of the segment "A" in Figure 2c would result in a power spectrum similar to that in Figure 2b, simply because the record length is not great enough to resolve the periodic fluctuation exhibited in Figure 2c.

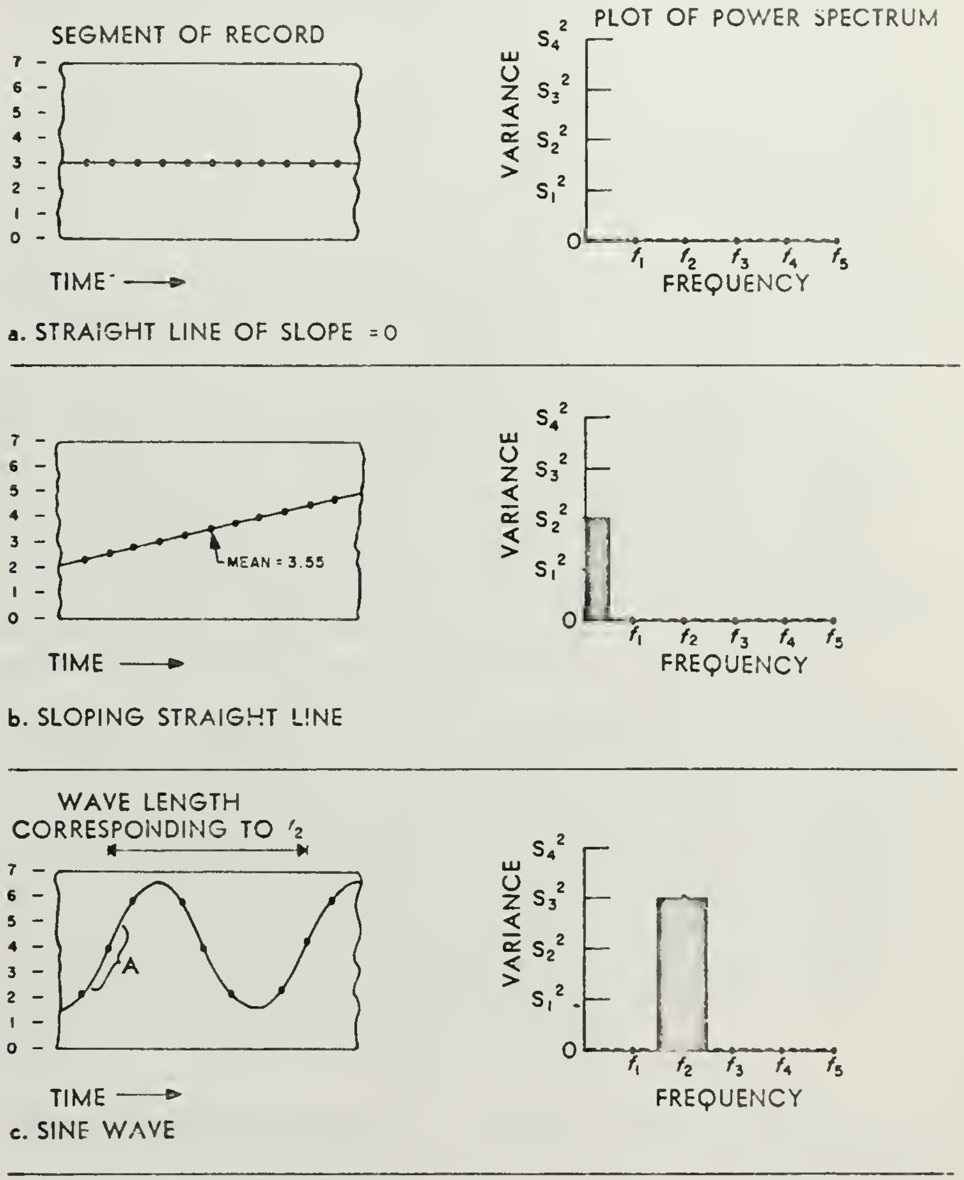


Figure 2. Typical spectra obtained from several types of curves.

The record of which Figure 2c is a small part also exhibits a variance. The mean of this record is 4.0, and the variance is entirely the result of the sinusoidal fluctuation about the mean. Spectral analysis of this record, which should be about 10 times longer than the segment presented in Figure 2c, would result in a power spectrum in which all of the variance is concentrated at the frequency designated f_2, which corresponds to the wave length of the sine wave in the data record.

If the data record were a combination of Figures 2b and 2c, the power spectrum would be a combination of the spectra in these figures, i.e., there would be components at the zero frequency and at f_2. It is this characteristic that makes spectral analysis such a useful tool in analyzing records that represent complex phenomena in natural systems. For example, the diurnal effects of photosynthetic activity could be separated from the longer-period effects of waste loads and river discharges by spectral analysis of the stream dissolved oxygen (DO) record.

The power spectra in Figures 1 and 2 are presented in bar graphs to emphasize two characteristics of spectral results: (1) estimates of the variance at several discrete frequencies are obtained from the analysis and (2) each variance estimate represents the variance concentrated in a band around the nominal frequency of each variance. This bar graph representation is not the usual way in which spectra are presented; in the remaining figures discussed in this paper, the conventional point-and-line representation is used.

Like any other method of data analysis, this method has its limitations and disadvantages. Three major requirements for the record may be regarded as limiting. First, the record must be fairly long -- generally having over 100 sequential measurements. Second, there must be no missing data -- if measurements are missing, suitable values must be interpolated before spectral analysis is attempted. Third, the mathematical procedures require so much computation that the use of a high-speed digital computer is essential for most analyses. These restrictions are discussed in more detail.

THE TECHNIQUE OF SPECTRAL ANALYSIS

The approach used here is to present the technique of spectral analysis by following through the steps in the actual spectral analysis of a record. For the benefit of those who wish to explore the mathematical basis for spectral analysis computations, pertinent references are presented in the bibliography.

all part also exhibits
.0, and the variance is
ation about the mean.
ould be about 10 times
ure 2c, would result in
riance is concentrated at
esponds to the wave
d.

on of Figures 2b and 2c,
tion of the spectra in
ponents at the zero fre-
ic that makes spectral
records that represent
For example, the di-
ould be separated from
and river discharges by
d oxygen (DO) record.

2 are presented in
stics of spectral results:
discrete frequencies
ach variance estimate
a band around the
is bar graph represen-
tr re presented; in the
er, e conventional

sis, this method has its
ajor requirements for
First, the record must
00 sequential measure-
ng data -- if measure-
t be interpolated before
the mathematical pro-
at the use of a high-
most analyses. These
il.

CTRAL ANALYSIS

nt the technique of
e steps in the actual
benefit of those who wish
ectral analysis compu-
nted in the bibliography.

The record chosen for detailed examination here is a record of water level obtained from a U. S. Geological Survey station in the Potomac Estuary near Washington, D. C. This record was chosen because it exhibits a simple periodicity with very little random interference ("noise"), and because the data were obtained as a continuous recording so that a wide choice of sampling intervals was possible.

A portion of this record is presented in Figure 3. A visual examination of this record shows that there is a dominant period of about 12 hours and that there is some long-period change.

Although the entire computation can be carried out on high-speed digital computers, with available programs, the individual steps are presented here to illustrate the technique. Only the initial steps in the data preparation need be carried out by hand or on semi-automatic equipment.

Step 1. A sampling interval of 4 hours was chosen for this particular analysis, and the record was read at this interval for a total of 145 readings. (Considerations governing the number of points read and the sampling interval are discussed later in detail.) The starting point was arbitrary. The number obtained, in order, are

	1.30	=	value 1
	2.57		2
	3.79		3
	1.49		4
	2.30		5
	4.73		6
	.		.
	.		.
	.		.
	.		.
	3.10		140
	1.46		141
	3.16		142
	3.30		143
	1.41		144
	2.35		145
Mean =	2.43		

Square of Mean = 5.90

Step 2. The autocorrelation function of these numbers is then formed. This is a very large name for a very simple, but very useful, process. Each number in the record is multiplied by another number in the record, and from the mean of the sum is subtracted the square of the arithmetic mean of the entire

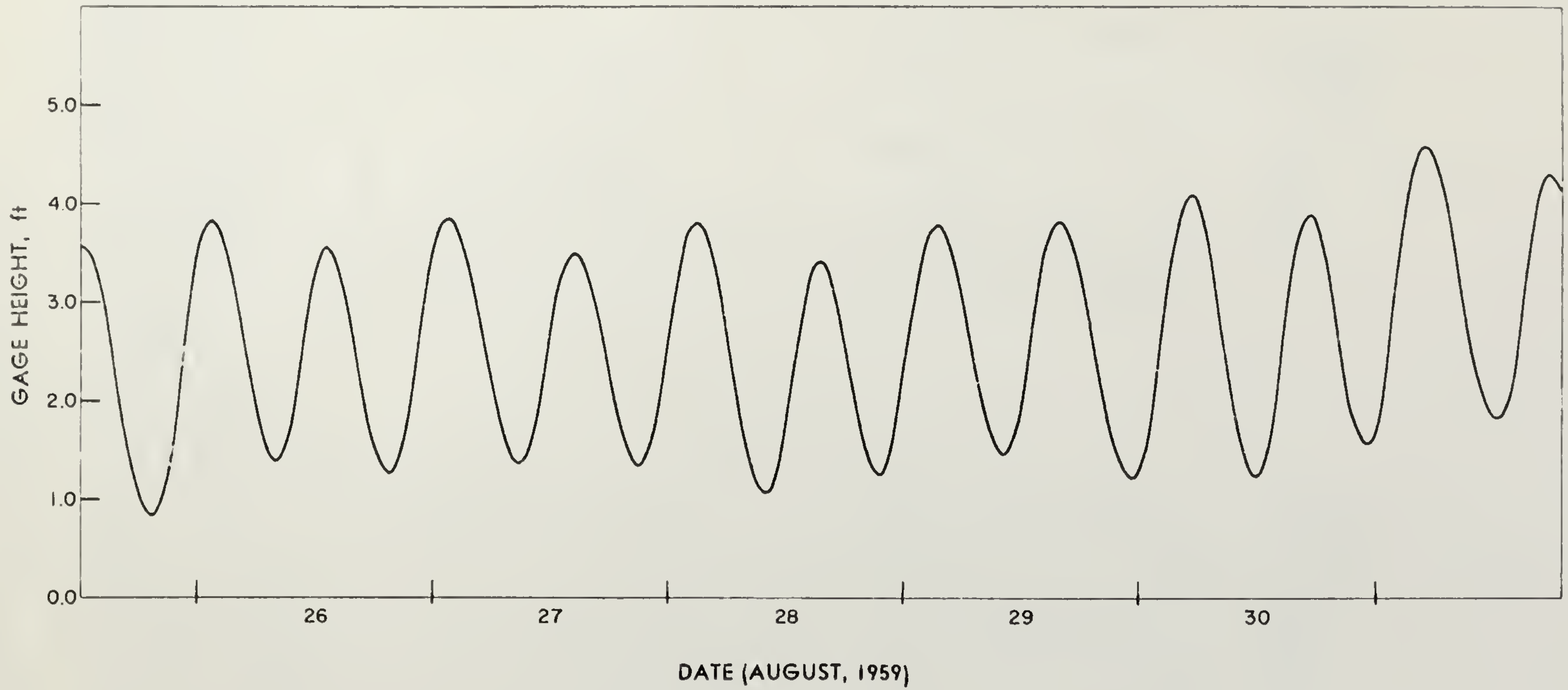


Figure 3. Portion of tide height record in the Potomac Estuary.

record. If the individual products are formed by multiplying each number by itself, the result is called the "autocorrelation at lag 0," which is purely and simply the "variance" as ordinarily defined in statistics. If the individual products are formed by multiplying each number by the number that follows it in the sequence, the result is called the "autocorrelation at lag 1." The autocorrelations computed to 12 lags for the record being analyzed here are

$C_0 \equiv$ Autocorrelation at lag 0	$C_1 \equiv$ Autocorrelation at lag 1
1.30 x 1.30 = 1.69	1.30 x 2.57 = 3.34
2.57 x 2.57 = 6.60	2.57 x 3.79 = 9.74
3.79 x 3.79 = 14.36	3.79 x 1.49 = 5.65
1.49 x 1.49 = 2.22	1.49 x 2.30 = 3.43
2.30 x 2.30 = 5.29	2.30 x 4.73 = 10.88
4.73 x 4.73 = 22.37	4.73 . .
. .	. . .
. .	. . .
. .	. . .
. .	. . .
3.10 x 3.10 = 9.61	3.10 x 1.46 = 4.53
1.46 x 1.46 = 2.13	1.46 x 3.16 = 4.61
3.16 x 3.16 = 9.99	3.16 x 3.30 = 10.43
3.30 x 3.30 - 10.89	3.30 x 1.41 = 4.65
1.41 x 1.41 = 1.99	1.41 x 2.35 = 3.31
2.35 x 2.35 = 5.82	2.35
Sum = 1046.9	Sum = 804.67

$$\frac{1046.9}{145} = 7.22 \qquad \frac{804.67}{144} = 5.588$$

$$\underline{-5.90} \qquad \underline{-5.90}$$

$$C_0 = 1.32 \qquad C_1 = -0.312$$

$C_2 \equiv$ Autocorrelation at lag 2

1.30 x 3.79 =	4.93
2.57 x 1.49 =	3.83
3.79 x 2.30 =	8.72
1.49 x 4.73 =	7.05
2.30 x .	.
4.73 x .	.
. .	.
. .	.
. .	.
. .	.
3.10 x 3.16 =	9.80
1.46 x 3.30 =	4.82
3.16 x 1.41 =	4.46
3.30 x 2.35 =	7.76
1.41	
2.35	

Sum = 764.62

$$\frac{764.62}{143} = 5.347$$
$$-5.90$$
$$C_2 = -0.553$$

$C_3 \equiv$ Autocorrelation at lag 3

1.30 x 1.49 =	1.94
2.57 x 2.30 =	5.91
3.79 x 4.73 =	17.93
1.49 x .	.
2.30 x .	.
4.73 x .	.
. .	.
. .	.
. .	.
. .	.
3.10 x 3.30 =	10.23
1.46 x 1.41 =	2.06
3.16 x 2.35 =	7.43
3.30	
1.41	
2.35	

Sum = 1015.3

$$\frac{1015.3}{142} = 7.10$$
$$-5.90$$
$$C_3 = 1.20$$

The remaining autocorrelations are computed similarly and have the values

$C_4 = 0.154$	$C_7 = 0.0476$	$C_{10} = 0.252$
$C_5 = 0.735$	$C_8 = -0.917$	$C_{11} = -0.998$
$C_6 = 1.10$	$C_9 = 0.921$	$C_{12} = 0.798$

This operation may be expressed mathematically as

$$C_r = \frac{1}{n-r} \sum_{1}^{n-r} x_t \, x_{t+r} - \left[\frac{1}{n} \sum_{1}^{n} x_t \right]^2 ,$$

where

C_r = autocorrelation at lag r,
x_t = record value at t,
t = 0, 1, 2 . . . n = sequential index of values,
r = 0, 1, 2 . . . m = lag number,
m = total number of lags.

correlation at lag 3

x 1.49 = 1.94

7 x 2.30 = 5.91

x 4.73 = 17.93

x . .

x . .

x . .

. .

. .

. .

. .

0 x 3.30 = 10.23

6 x 1.41 = 2.06

6 x 2.35 = 7.43

30

41

35

um = 1015.3

$$\frac{1015.3}{142} = 7.10$$

$$-5.90$$

$$C_3 = 1.20$$

puted similarly and have

$$C_{10} = 0.252$$
$$C_{11} = -0.998$$
$$C_{12} = 0.798$$

atically as

$$\sum_{1}^{n} x_t \Big]^2 ,$$

x of values,

Step 3. The Fourier cosine transform for each autocorrelation is next computed. This serves to smooth out some of the wide fluctuations in the autocorrelations and consists of applying a cosinusoidal weighting factor to each autocorrelation calculated in the preceding step. This operation can be expressed mathematically as

$$V_r = \frac{k}{m}\left[C_0 + C_m \cos r\pi + 2\sum_{q=1}^{m-1} C_q \cos \frac{q r \pi}{m}\right],$$

where

V_r = Fourier cosine transform of the autocorrelation at lag r,

q = lag number, having values between 1 and m-1

k = a constant, $k = 1$ for $r = 1, 2 \ldots m-1$,
$k = 1/2$ for $r = 0$,
$r = m$,

and the other letters have the definitions previously given.

The Fourier cosine transforms calculated for each autocorrelation of the tide height record are

$$V_0 = \frac{1}{(2)(12)}\left[1.32 + 2\left(\sum_{q=1}^{m-1} C_q \cdot \cos \frac{q(0)\pi}{m}\right) + 0.798 \cos(0)\pi\right]$$

$$= \frac{1}{24}\Big[1.32 + 2(1)(-.312 - .553 + 1.20 - 0.154 - 0.735 + 1.10 + 0.0476 - 0.917 + 0.921 + 0.252 - 0.998) + 0.798\,(1)\Big]$$

$$= 0.0765$$

$$V_1 = \frac{1}{12}\Big[1.32 + (0.798)\cos(1)\pi\Big] + \frac{2}{12}\Big[-(0.312)\cos\frac{(1)(1)\pi}{12} - (0.553)\cos\frac{(2)(1)\pi}{12} + (1.20)\cos\frac{(3)(1)\pi}{12} - (0.154)\cos\frac{(4)(1)\pi}{12} - (0.735)\cos\frac{(5)(1)\pi}{12} + (1.10)\cos\frac{(6)(1)\pi}{12} + (0.0476)\cos\frac{(7)(1)\pi}{12} - (0.917)\cos\frac{(8)(1)\pi}{12} + (0.921)\cos\frac{(9)(1)\pi}{12} + (0.252)\cos\frac{(10)(1)\pi}{12} - (0.998)\cos\frac{(11)(1)\pi}{12}\Big]$$

$$V_1 = 0.101$$

Similarly, the Fourier cosine transforms for the remaining autocorrelations can be computed:

$V_2 = -0.0332$	$V_8 = 0.995$
$V_3 = 0.0570$	$V_9 = -0.224$
$V_4 = -0.0443$	$V_{10} = 0.146$
$V_5 = 0.0943$	$V_{11} = -0.118$
$V_6 = -0.136$	$V_{12} = 0.0561$
$V_7 = 0.353$	

Step 4. The final step in the spectral analysis of a single record is another weighting operation that counteracts some distortion of the spectrum resulting from the small sample size.

This step can be expressed mathematically as

$$U_0 = 0.54 \left[V_0 + V_1 \right],$$

$$U_r = 0.23 \; V_{r-1} + 0.54 \; V_r + 0.23 \; V_{r+1} \,,$$

for $r = 1, 2, 3 \ldots\ldots m-1$

$$U_m = 0.54 \; V_{m-1} + 0.54 \; V_m \,,$$

where U_0, U_r, U_m are the power spectrum estimates corresponding to the respective lags, and the remaining symbols have the meanings previously assigned.

The power spectrum estimates for the tide gage record are

$U_0 = (0.54)(0.0765) + (0.54)(0.101) = 0.0959$

$U_1 = 0.0643$	
$U_2 = 0.0183$	
$U_3 = 0.0130$	$U_8 = 0.567$
$U_4 = 0.0109$	$U_9 = 0.142$
$U_5 = 0.00947$	$U_{10} = 0.000446$
$U_6 = 0.0295$	$U_{11} = -0.0173$
$U_7 = 0.388$	$U_{12} = -0.0334$

Each of these spectral estimates represents the part of the total record variance that is estimated to occur with a certain periodicity. The period corresponding to each lag is determined from the lag number and the sampling interval by this relation:

$$T_r = \frac{2m \quad \Delta \tau}{r} \,,$$

where T_r = period corresponding to lag r,

$\Delta \tau$ = sampling interval.

.995

.224

.146

.118

.0561

alysis of a single record
eracts some distortion
sample size.

tically as

V_{r+1} ,

um estimates corres-
emaining symbols have

he tide gage record are
0.0959

).567

).142

).000446

).0173

).0334

presents the part of the
o occur with a certain
o each lag is determined
terval by this relation:

lag r,

For the record analyzed here, m = 12 and $\Delta\tau$ = 4 hours. The periods corresponding to the lag numbers are these:

Lag Number	Period (hours)
0	∞
1	96
2	48
3	32
4	24
5	19.2
6	16
7	13.7
8	12
9	10.7
10	9.6
11	8.7
12	8

The spectral estimates are plotted as functions of period in Figure 4.

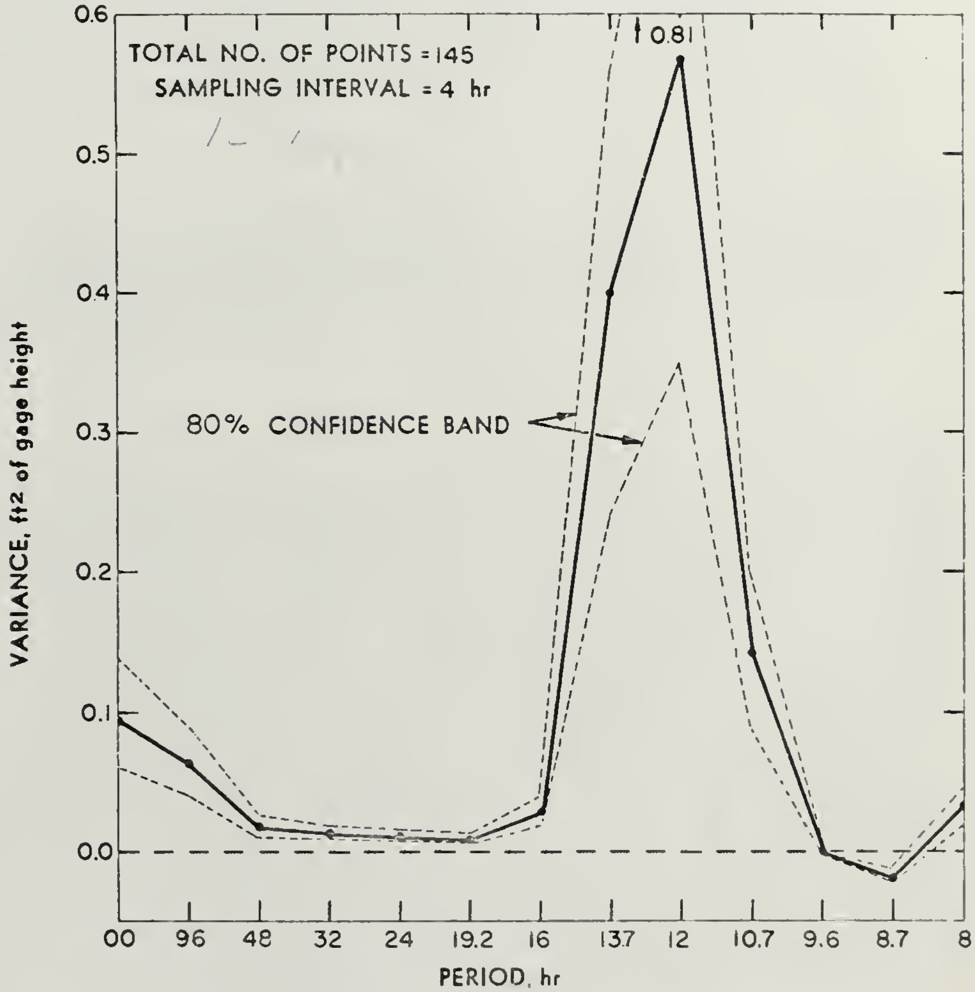


Figure 4. Spectrum of a tidal height record.

From these results it can be seen that about 79 percent of the observed variance in water level can be attributed to periodicities of around 12 hours, that there are very small diurnal effects in the tide at this point, and that long-period changes account for about 12 percent of the observed variance.

The effects of using different sampling intervals and computing to different numbers of lags may be examined by considering the results of further spectral computation with this record. Figures 4, 5, and 6 illustrate spectra obtained from the same record, with variations only in the sampling interval and in the number of lags used in calculation. The total record length in each case was 576 hours. The spectral estimates from which Figure 4 was plotted were obtained by computation to 12 lags from values read at 4-hour intervals (145 points). Figure 5 was obtained from computation to 12 lags from values read at hourly intervals (577 points). Figure 6 was obtained from computation to 24 lags from values read at hourly intervals. The abscissal scale in Figure 6b is 50 times greater than that in Figure 6a; this shows short-period effects more clearly.

While each of these figures shows the dominance of approximately semidiurnal periodicities in this record, the effects of the different sampling intervals and the different numbers of lags can be seen also.

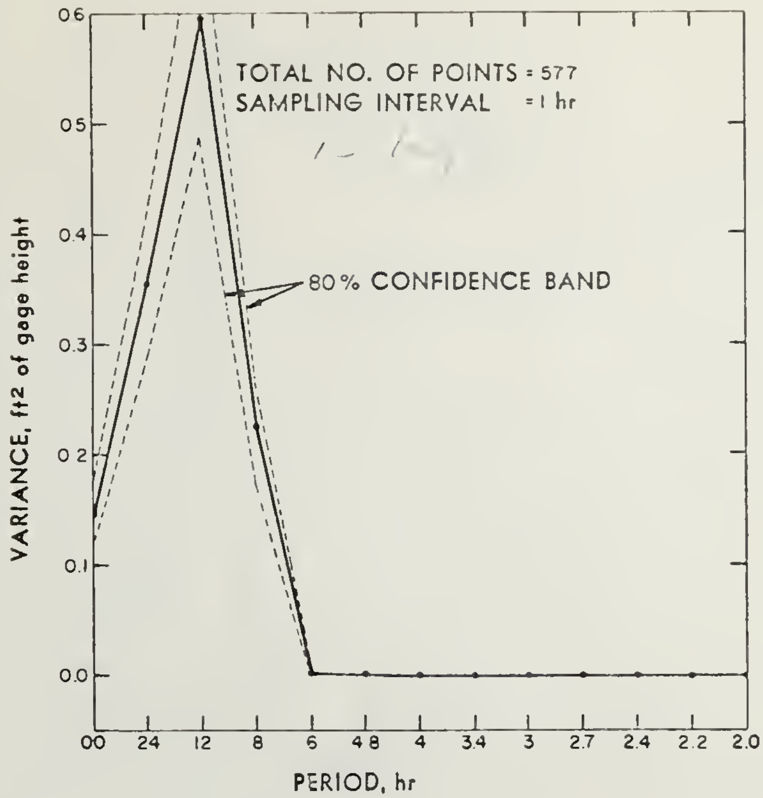
Comparison of Figures 4 and 5 shows the effects of a change in sampling interval on the spectrum of the tide height record. First, the use of a smaller sampling interval for a given record length increases the number of measurements used in the analysis and therefore increases the number of degrees of freedom upon which each estimate is based. This results in the smaller confidence band shown in Figure 5. Second, the change of sampling interval from 4 hours to 1 hour permits the resolution of components with periods as short as 2 hours in the latter case instead the 8 hours possible in the former. However, the use of the same number of lags with the shorter sampling interval does not permit as high a degree of resolution of long-period phenomena as was obtained with the longer sampling interval. In fact, the spectrum calculation leading to Figure 5 does not permit an estimate of the diurnal and longer-period effects because these are effectively masked by the dominant 12-hour component.

about 79 percent of
attributed to periodi-
ry small diurnal
g-period changes
d variance.

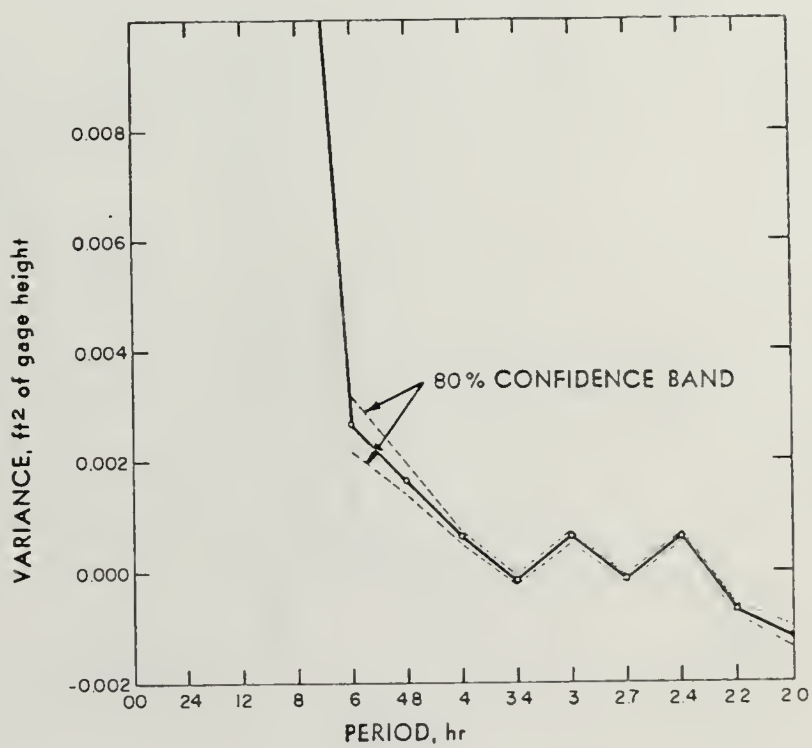
intervals and com-
examined by consider-
tion with this record.
ined from the same
ig interval and in the
otal record length in
timates from which
putation to 12 lags
points). Figure 5
rom values read at
s obtained from com-
rly intervals. The
reater than that in
more clearly.

dominance of approxi-
ecord, the effects of
fferent numbers of

s the effects of a change
he tide height record.
rval for a given record
ents used in the analy-
f degrees of freedom
results in the smaller
nd, the change of sam-
nits the resolution of
urs in the latter case
r. However, the use of
sampling interval does
of long-period pheno-
npling interval. In fact,
e 5 does not permit an
effects because these
2-hour component.

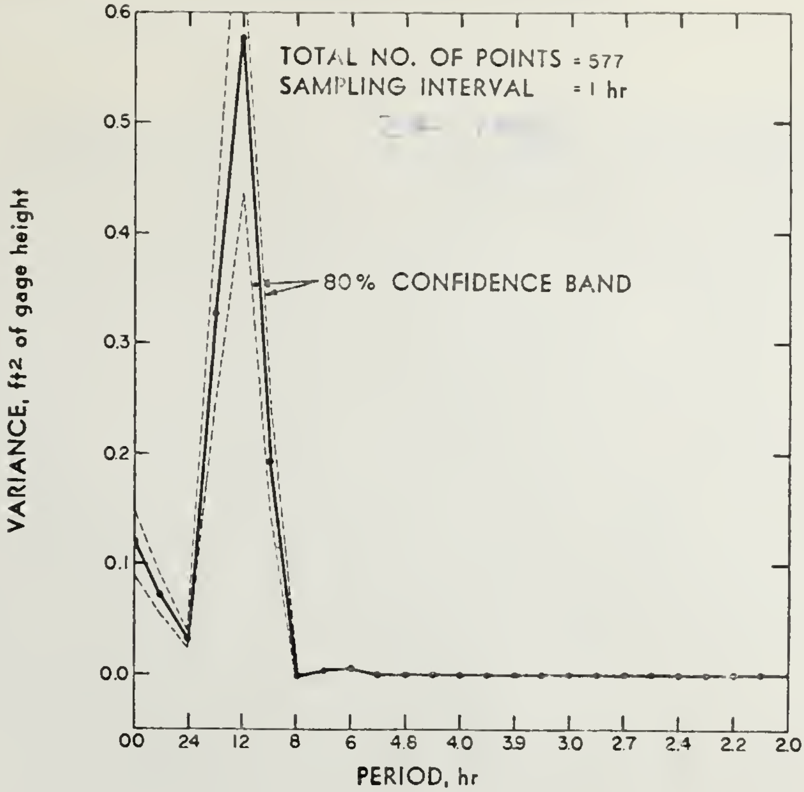


5a

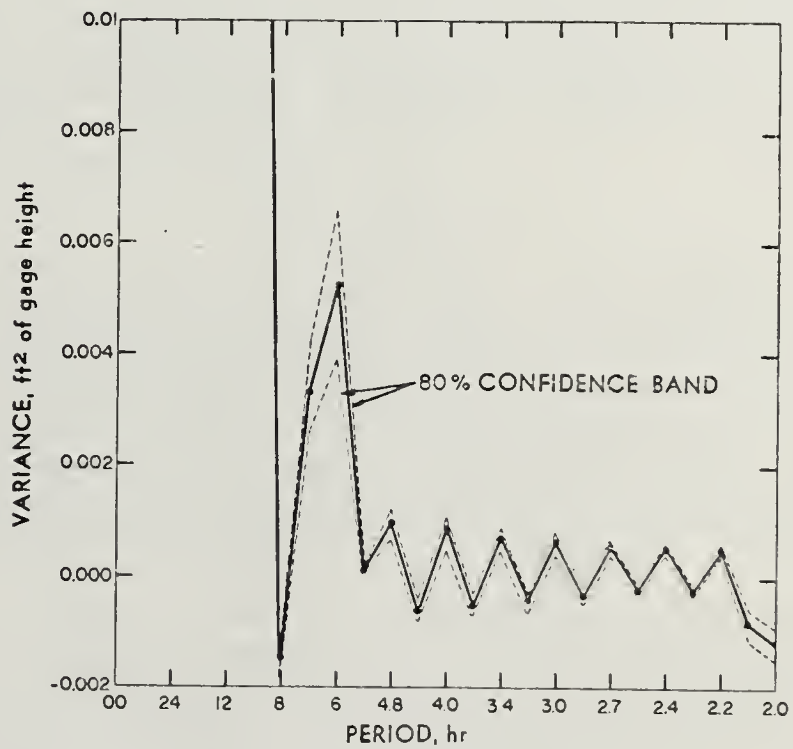


5b

Figure 5. Spectrum of a tidal height record.

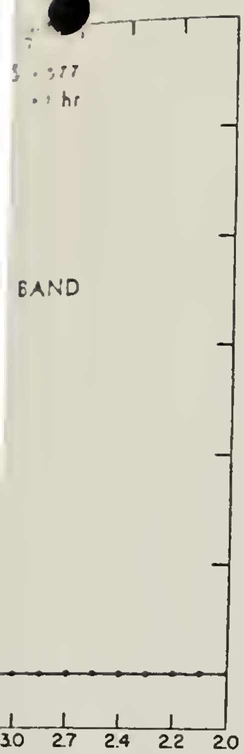


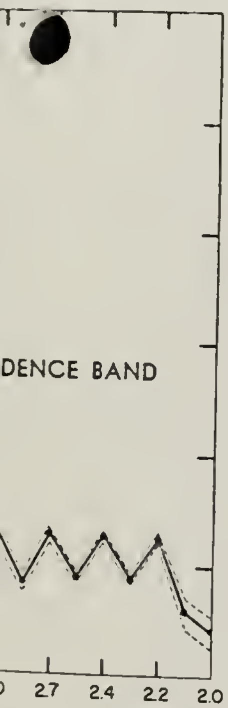
6a



6b

Figure 6. Spectrum of a tidal height record.





t record.

Comparison of Figure 5 with Figure 6 shows the effect of increasing the number of lags without changing the sampling interval. The resolving power is increased, i.e., estimates are made for a greater number of spectral bands, but with a slight loss of confidence in the estimates, as is indicated by the wider confidence band in Figure 6 than in Figure 5. In this particular case, however, the increased resolving power demonstrates the existence of a small but statistically significant overtide (a harmonic of the semidiurnal tidal component) with a period of about 6 hours. This overtide is not shown in Figure 5b because the dominating semidiurnal band has overlapped the 6-hour-period band sufficiently to mask the very small overtide. In Figure 6b the resolving power of the 24 lags used in computation produces band-widths sufficiently narrow to prevent overlap of the 12- and 6-hour periods. The existence of higher harmonics is also shown in both Figure 5b and Figure 6b. The former figure shows significant fourth and fifth harmonic overtides, whereas the latter shows these and several other short-period components.

The results of the spectral analysis of this record may be used to point out several characteristics of this technique.

First, each of the spectral values obtained represents an estimate of the variance over a range of periods in the vicinity of the nominal value. The range of periods for which each estimate is computed is determined by the sampling interval and the number of lags used in computation. This range is called the "equivalent width" (W_e) of the spectral band and can be calculated in terms of frequency by

$$W_e = \frac{1}{2\, m\, \Delta\tau}$$

The spectral estimate reported is an average value for all periods in the band over the range W_e. For example, for the results presented in Figure 4, the spectral estimate reported as corresponding to a 12-hour period is actually an average for the range of periods from 10.7 hours to 13.7 hours. The overlapping of spectral estimates is illustrated in Figure 5, and Figure 6 demonstrates how this overlapping can be reduced by increasing the number of lags, thereby reducing the equivalent width of the band for each spectral estimate.

Second, the precision of each estimate is a function of the total number of samples and the number of lags used in computation. A method for estimating the number of degrees of freedom for each estimate and for establishing the confidence intervals has been presented by Blackman and Tukey[3]. In this method, the process being measured is regarded as Gaussian and the degrees of freedom and confidence intervals are based on a Chi-square distribution. The 80 percent confidence bands

indicated on Figures 4, 5, and 6 have been determined by this method. It has been previously noted that a decrease in the total number of samples or an increase in the number of lags causes a widening of the confidence band, which represents a loss of precision in each spectral estimate.

Third, if the sampling interval is not small enough to permit resolution of the shortest periods that contribute significant variance to the record, the short-period variance in the record is not lost but is reported at harmonics of the true period. For example, if this record, as analyzed in Figure 4, contained significant variance with a period of 4 hours, this would be shown in the computations as part of the spectral estimate for the 8-hour period. This occurrence is known as "aliasing" or "folding." In Figure 4, the appearance of a significantly large estimate at the shortest period computed in the analysis indicates that there is probably some aliasing of the short-period variance. The use of a smaller sampling interval, as shown in Figure 5, eliminated the aliasing in this spectrum.

Fourth, the appearance of an occasional negative spectral estimate is an artifact of the computation and results from the use of a finite record length for spectral computations. If the record were infinitely long, there would be no negative spectral estimates; but, when an estimate is close to zero in magnitude, it may appear with either a positive or negative sign. This result is interpreted only as a very small quantity of variance, and no physical significance is attributed to the negative sign.

INTERPRETATION OF SPECTRA

In order to illustrate how the computed spectra of a set of observations can be used to gain an insight into the structure of a river or estuarine system, some data obtained in a field survey of the Potomac Estuary near Washington, D.C., are presented here as a basis for discussion. Since the purpose of this discussion is to present a conceptual picture of how spectral results may be interpreted, the quantitative results obtained from this particular study are not presented at this time.

In Figures 7 and 8 are presented time series of DO and 5-day biochemical oxygen demand (BOD) measurements obtained at six stations in this estuary. In Figures 9 and 10 are the spectra computed from these records. Each record contained 145 points obtained at 4-hour intervals; computation was carried out to 12 lags. Figure 11 is a schematic representation of the station locations.

determined by this
a decrease in the total
number of lags causes
represents a loss of

small enough to per-
contribute signifi-
od variance in the
nics of the true period.
in Figure 4, contained
rs, this would be
pectral estimate for
wn as "aliasing" or
a significantly large
in the analysis indi-
of the short-period
interval, as shown in
pectrum.

al negative spectral
and results from the
omputations. If the
r no negative spectral
to zero in magnitude,
sign. This
quantity of variance,
to the negative sign.

SPECTRA

spectra of a set of
into the structure of
tained in a field sur-
n, D.C., are pre-
the purpose of
cture of how spec-
tive results obtained
at this time.

ries of DO and
asurements ob-
res 9 and 10 are
Each record con-
computation was
atic representation

Each spectrum shown in Figures 9 and 10 consists of three major effects: long-period, diurnal (24-hour), and semidiurnal (12-hour). The estimates adjacent to the long-period, 24-hour, and 12-hour periods are affected by the strength of the dominating period and the overlapping of spectral bands, as was discussed earlier in relation to the tide gage record. The significance of each of these effects may be considered separately.

The effects reported in the computation as "long-period" or of "infinite" period include all components whose periods are too long to be resolved in the computation. Since in these computations there is overlapping of the 96-hour and "infinite" periods, this means that all components with periods longer than 48 hours are regarded as long-period components. The use of the infinity symbol in the figures is merely a convenience in plotting and should be interpreted as meaning "long-period."

The components reported as long-period may have one or more of these physical interpretations:

1. The record may be affected by the existence of periodic components that are too long to be resolved by the length of record available. For example, a record of air temperature for a length of several months might show a long-period effect that, from a record of several years duration, would appear as an annual cycle.
2. There may be a secular trend in the record that is fundamentally aperiodic in nature. For example, a record of world population over several centuries might show such a trend.
3. Random sampling, reading, and analytical errors appear as long-period effects. A constant bias in the data affects the mean of the record only, and since spectra are computed from deviations from the mean, such a bias would not affect the spectral results.

From the computed spectra alone, it is not possible to differentiate among random errors, aperiodic effects, and long-period effects. In a practical sense it is often quite adequate to regard the long-period spectral estimates as secular trends for the available record length and to regard any random error as constant for all records of the same measurement.

The DO spectra presented in Figure 9 exhibit long-period effects that generally decrease from the upstream to the downstream stations. If a constant random error is assumed, the existence of an effect that is a function of distance from the head of tidewater may be postulated, and it might be deduced that this effect is related to the river discharge entering the estuary above Station 3. Since the parameter being measured

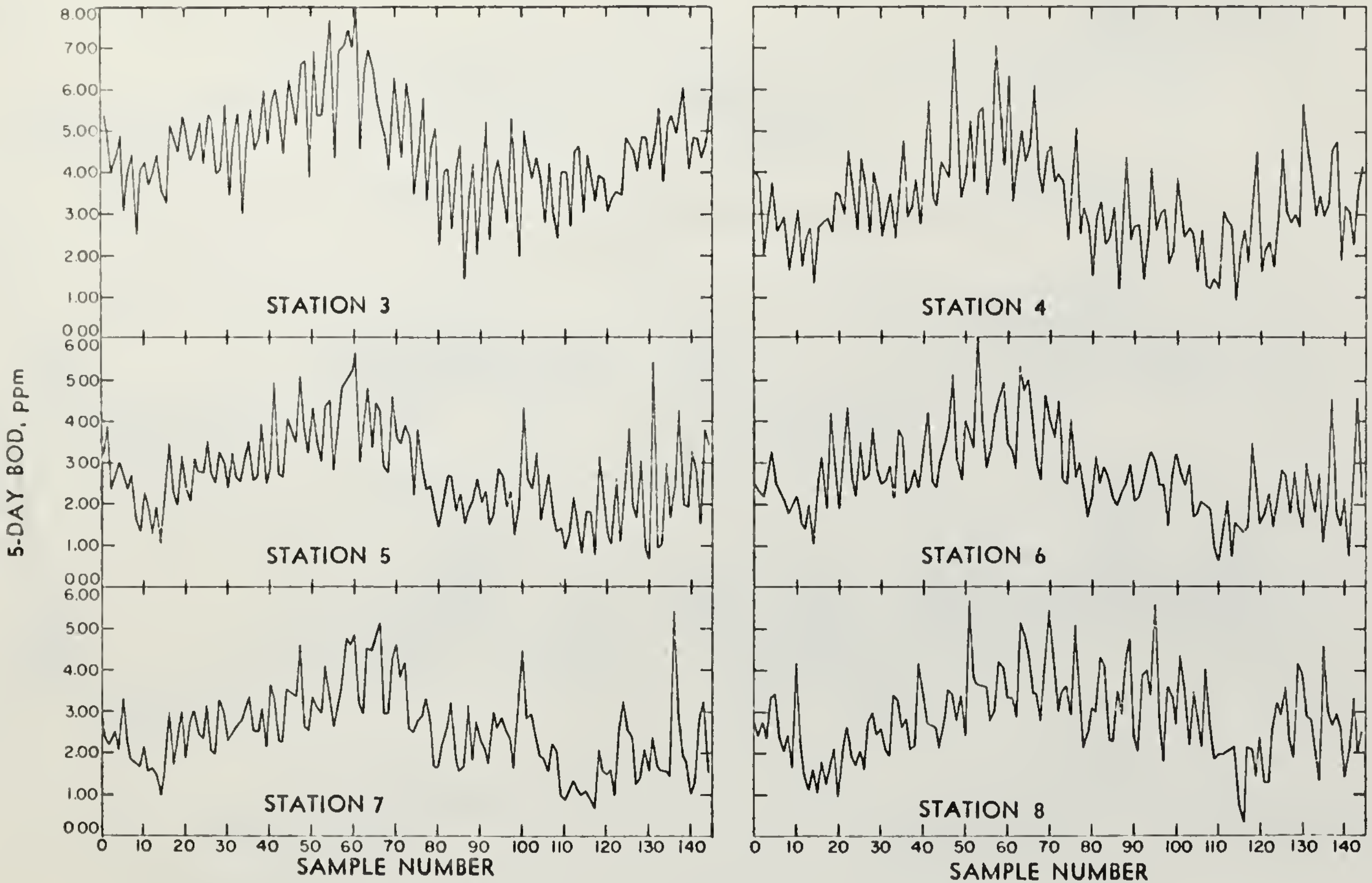


Figure 7. Dissolved oxygen records obtained in the Potomac Estuary, August 1959.

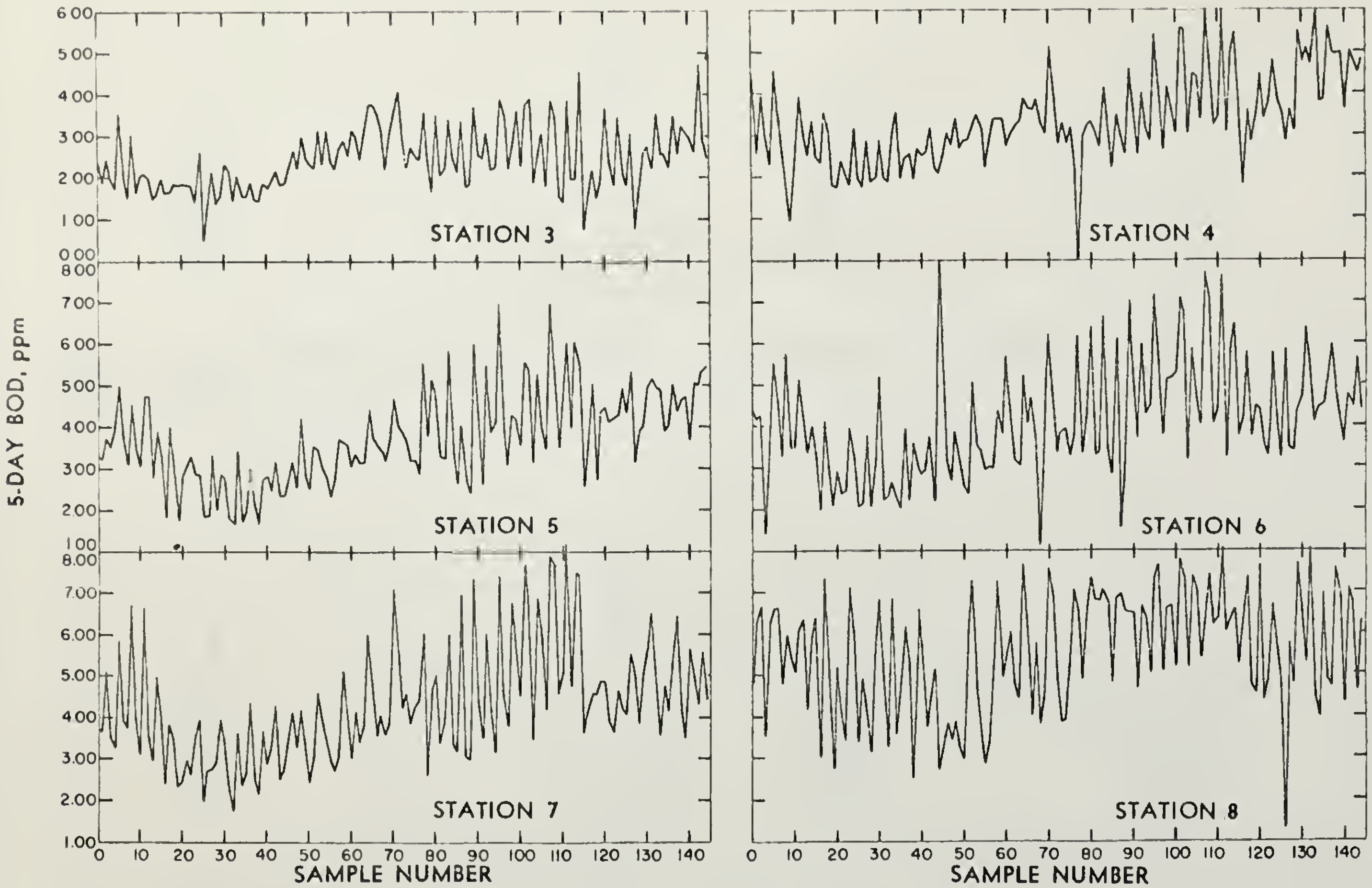
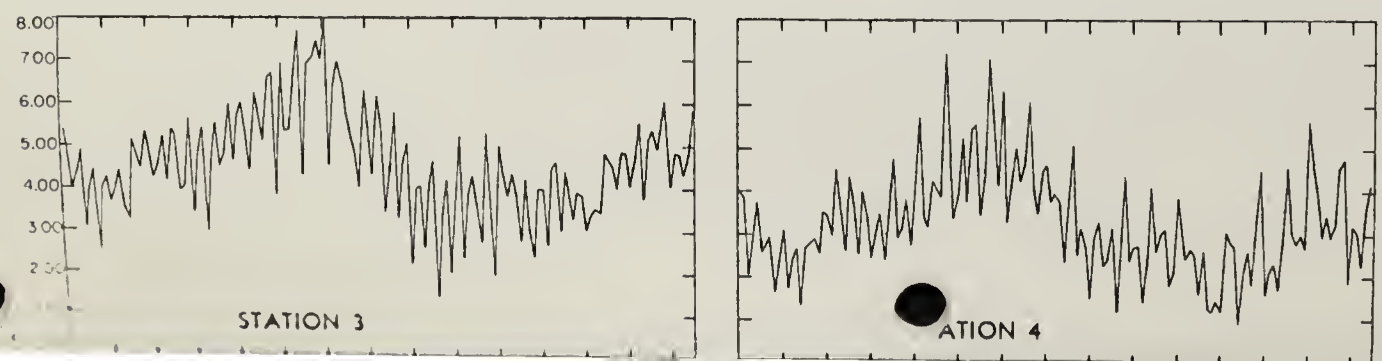


Figure 8. Biochemical oxygen demand records obtained in the Potomac Estuary, August 1959.



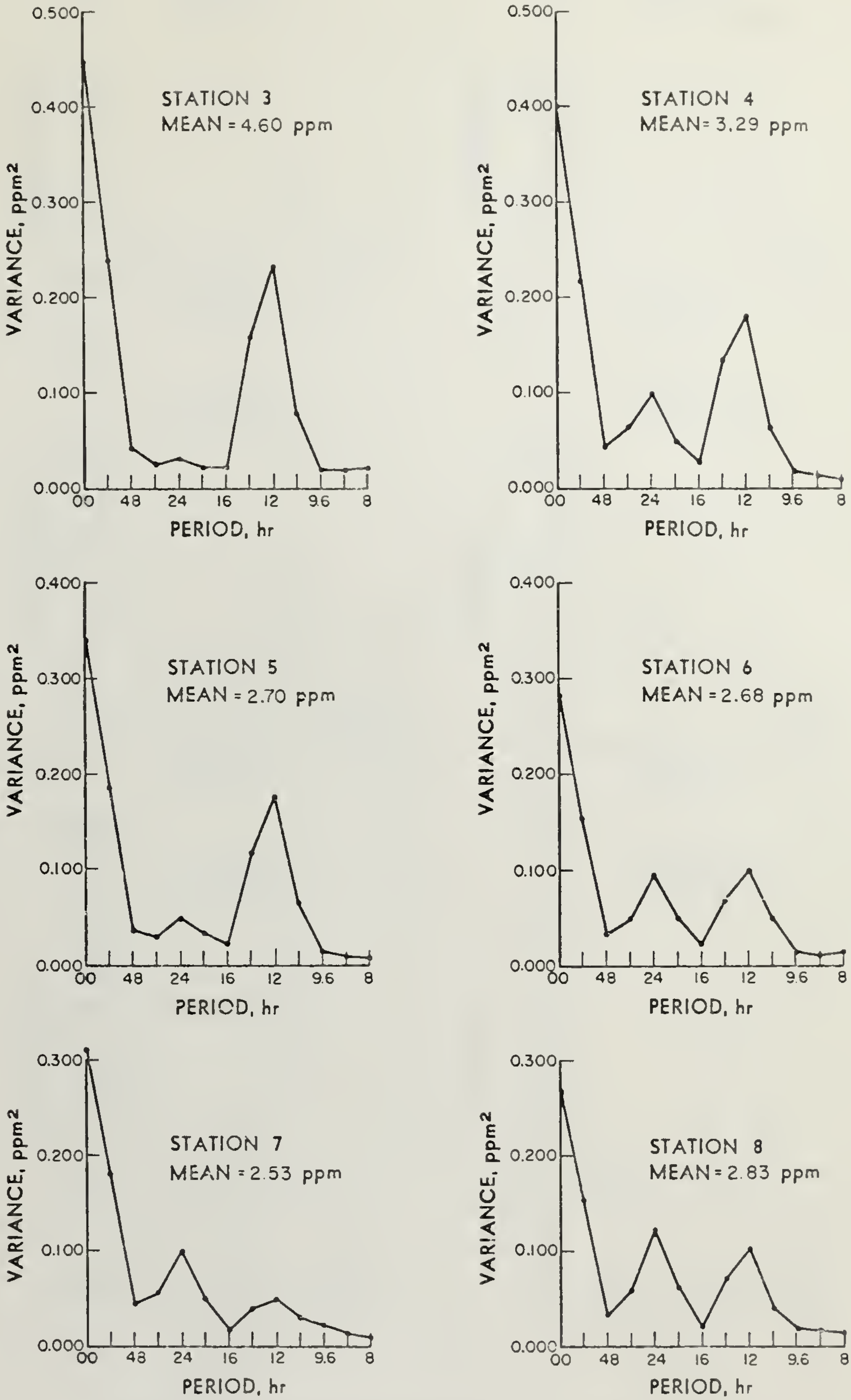
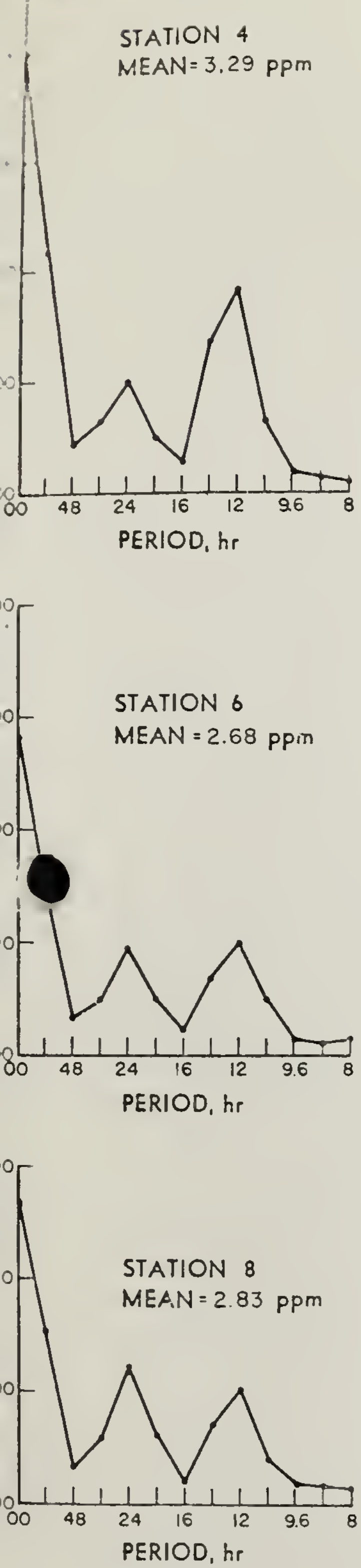


Figure 9. Dissolved oxygen spectra in the Potomac Estuary.



Potomac Estuary.

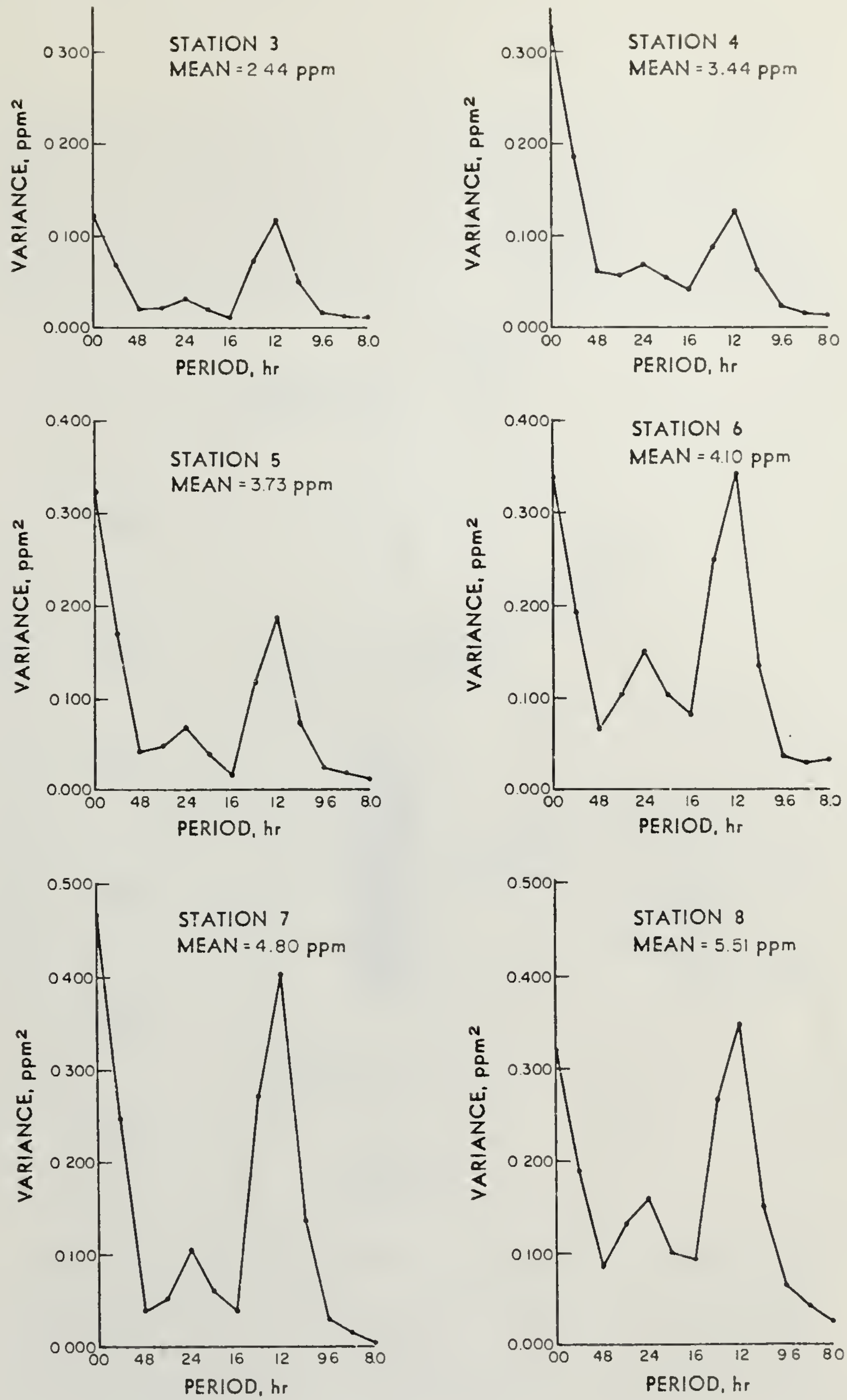


Figure 10. Biochemical oxygen demand spectra in the Potomac Estuary.

is DO this long-period effect might be interpreted as a measure of the amount of DO advected to the system in the river discharge or as an effect of river discharge on the reaeration characteristics of the system, or both. Considerable caution is in order if these results are to be interpreted in this manner; the change in the long-period estimate from Station 3 to Station 8 is barely significant at the 80 percent level, and the differences

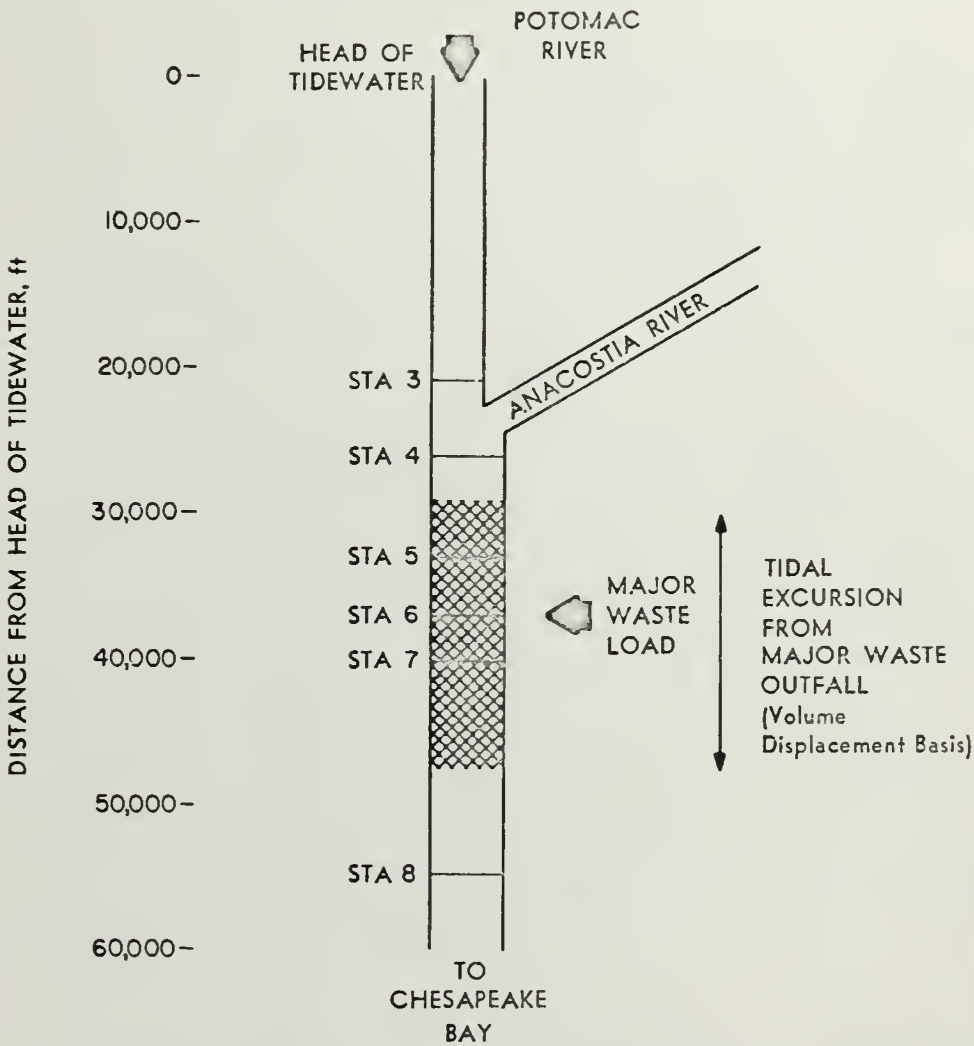


Figure 11. Schematic representation of Potomac Estuary sampling station locations.

in this estimate between successive stations are not significant at the 80 percent level. The existence of a definite trend does indicate, however, that the interpretation given is reasonable and affords a basis for the examination of other results in terms of this hypothesis.

The BOD spectra in Figure 10 present an aspect somewhat different from the DO spectra at the same stations. The long-period effects at four of the six stations have very similar values.

Station 7 exhibits a long-period component that is somewhat higher than these but still lies within the 80 percent confidence band of the four stations. Station 3 exhibits a long-period effect that is significantly less than the corresponding effects at the other five stations. These components at the five downstream stations can be interpreted as the result of a secular trend in the major waste discharge to the system as well as random error in the determination of BOD.

Comparison of the long-period effects at Stations 3 and 4 with those of stations in the immediate vicinity of the waste outfall suggests that the limits of a tidal excursion upstream from the outfall may actually lie between Stations 3 and 4 instead of downstream of Station 4, as the calculations based on the tidal prism volume displacement indicate. The long-period effects at Station 3 might then be regarded as resulting from the longitudinal mixing of the waste discharge, whereas the similar effects at the other stations may be regarded as reflecting a combination of advective and diffusive processes.

The diurnal components of the DO and BOD spectra may be interpreted as expressing the effects of diurnal variations in waste discharge and in photosynthetic activity of the planktonic population. Diurnal variations in waste discharge would affect stations within a tidal excursion of the outfall more strongly than those beyond this distance, whereas photosynthetic activity in the system would be more pronounced at those stations exhibiting the higher nutrient concentrations, generally reflected in higher mean stream BOD's. Interpreted on this basis, the diurnal spectral components suggest that the extent of the upstream tidal excursion is between Stations 3 and 4, whereas Stations 6, 7, and 8 are subject to considerable photosynthetic activity in addition to the diurnal waste load variations. The diurnal components of the DO spectra at each station correspond in size to the respective BOD components; this is a result that might be expected from the theory of DO-BOD relationships in streams.

The semidiurnal component of these spectra reflects the advective motion of DO and stream waste load due to tidal action. The magnitude of this component at each station is a measure of the longitudinal concentration gradient of each parameter within a tidal excursion of the station. The DO spectra show large semidiurnal effects at Stations 3, 4, and 5, and relatively smaller ones at Stations 6, 7, and 8; these results indicate that there is a large DO gradient in the upper part of the system and a small one in the lower part. Examination of the mean DO values at each station suggests the existence of an oxygen-sag regime in which Stations 3 and 4 represent a zone of rapid degradation and Stations 5, 6, 7, and 8 a zone of critical DO and the beginning of

recovery, a situation that is in close agreement with the spectral results. From this point of view the region of critical DO is the location that has the smallest semidiurnal spectral component, in this case Station 7.

The semidiurnal BOD spectral components present a picture somewhat different than the DO spectra. At the two upstream stations this component of the BOD spectrum is small. There is a significant increase in this component at Station 5 and again at Station 6, whereas the semidiurnal effects at Stations 7 and 8 are of the same magnitude as that at Station 6. These results suggest a low BOD gradient in the upper reaches of the system and a high gradient in the lower reaches, and Station 5 represents a region of transition. The strong gradients near and below the waste outfall suggest that the semidiuranl spectral components are affected by variations in the initial mixing of the waste load throughout the volume of water passing the outfall in a tidal excursion. It is apparent that the mean BOD's do not show such changes in gradient along the estuary. Comparison of the BOD gradient regime shown by the spectral analysis (with that shown by the mean BOD values at each station illustrates the sensitivity of spectral analysis) as a tool in estuarine engineering.

It has been the purpose of this discussion to demonstrate how spectral results can be interpreted in terms of familiar sanitary engineering concepts. It is not intended to suggest that the spectra can supply no information in addition to that discussed. Spectral analysis is purely and simply a tool for the analysis of time-series data. It provides by itself no theoretican insight into natural processes, but it does permit one to examine individual periodic components of the data, with a minimum of interference from other components. As with any other statistical technique, the final interpretation of these results must be based on an understanding of the natural process, not on some magic numbers produced by the manipulation of data.

DESIGN CRITERIA FOR SPECTRAL ANALYSIS

In the design of any spectral analysis program the requirements for precision in each spectral estimate, for resolution of sufficient spectral bands, and for eliminating aliasing, or folding, must all be considered and balanced against each other before the sampling program and data analysis are begun.

An acceptable balance among these requirements can be established by careful choice of the sampling interval, the record length, and the number of lags used in computation.

The choice of appropriate sampling and computational factors should be based on reasonable assumptions of what the dominant periodicities in the system are. The shortest period it is necessary to resolve determines the sampling interval required, whereas the longest period necessary determines the total record length. In most sanitary engineering applications, it may be assumed that diurnal fluctuations in waste discharges and in photosynthetic activity will be of considerable importance. In tidal systems there will be significant semidiurnal periodicities; there may also be some effects of this short a period in waste discharges and in the biological regime for a non tidal system. As a general basis for experimental design, it may be assumed that resolution of 24-hour periods will be required in a non tidal system and that resolution of 12-hour periods will be required in a tidal system.

The shortest period it is theoretically possible to resolve with a given sampling interval is the period that is twice the sampling interval. That is, with a sampling interval of 6 hours, it is theoretically possible to resolve a 12-hour period. From a practical standpoint it is not possible to do this, since the 12-hour estimate would be the shortest period computed, and there would arise the question whether this is a valid estimate of a 12-hour period or whether it merely represents the aliasing of periods shorter than 12 hours. It is advisable that the sampling interval chosen be small enough to provide spectral estimates for several periods shorter than the expected dominant shortest period, so that this period is minimally affected by any aliasing that might occur.

As a rule of thumb, it is suggested that a maximum sampling interval of 8 hours is required to resolve 24-hour periods and an interval of 4 hours to resolve 12-hour periods. In general, a sampling interval of not more than one-third the length of the shortest significant period is recommended.

There are no clear criteria for determining the longest period that can be resolved from a given record. The longest period resolved in any particular analysis (other than the "infinite" period estimate) is determined by the number of lags used in computation and by the sampling interval. In general, computation to a number of lags greater than 10 percent of the total number of measurements in the record is not recommended, i.e., for a record of 140 measurements, computation to no more than 14 lags is recommended. Each additional lag used in computation reduces the precision of all spectral estimates computed, and it is generally regarded that the 10 percent value affords an optimum balance between precision of individual estimates and resolution of spectral components. As a guide in determining the required record length for design purposes, it may be assumed that a record length at least 10 times as long as

the longest significant period to be resolved will be required. For many sanitary engineering field surveys, the 24-hour period is about the longest significant period it is necessary to resolve from field survey data; in such cases a minimum record length of 240 hours would be required.

As a basis for selecting the number of lags to be used in computation 10 percent of the number of measurements in the record is used as an upper limit. The number of lags finally chosen will probably be based on the resolution believed to be required. For example, if it is desired to separate diurnal and semidiurnal components, the number of lags chosen must be large enough to include several estimates between the 24-hour and 12-hour estimates, so that there is a negligible amount of interference between the major components. An acceptable degree of precision is then obtained by increasing the record length if the numbers of degrees of freedom on which each estimate is based give a confidence band that is too large to give usable results.

Any type of time-series record can be subjected to spectral analysis if it represents sampling at uniform time intervals and if there are no missing points; if a few points are missing, however, a limited amount of interpolation may be done. Interpolation of the mean value of all the measurements or linear interpolation of a missing point between two measurements is the usual approach. There are no definite criteria that serve as an indication of how much interpolation can be done in any particular case; a general consideration of the process suggests, however, that if the missing points are widely scattered up to possibly 5 per cent of the data may be interpolated without serious effects on the computed spectra.

The obtaining of data suitable for spectral analysis is most simply and cheaply accomplished by means of automatic sampling and recording equipment. Conversely, spectral analysis offers the most effective means of analyzing and correlating the large quantities of data produced by such instruments. This does not mean, of course, that the computation of spectra is limited to time-series data obtained from automatic devices. The DO and BOD data discussed here were obtained by conventional manual sampling procedures, whereas the tide height record was obtained from an automatic recording device. If the requirements of uniform sampling interval and record length are met, the means by which the data are obtained is immaterial.

The computation of spectra is most cheaply and efficiently accomplished by high-speed digital computer. The spectra presented here were computed on an IBM 704; the total cost, including preparation of the data for the computer, was estimated

at less than 10 dollars per spectrum. Computation by hand for all but very short record lengths is prohibitive from the standpoints of time, money, and accuracy.

CONCLUSION

It has been the purpose of this paper to present the technique of spectral analysis as a statistical tool that can be used in a wide variety of sanitary engineering applications. A strictly operational viewpoint has been maintained, and the theoretical basis of generalized harmonic analysis and spectral computation has been ignored.

It is unfortunate that there is no single work that can be offered as a primary reference on the theory of spectral analysis. Pertinent references on the theory and practice of spectral analysis are given in the bibliography. Of these, Panofsky and Briar present an introduction to the subject with emphasis on the meteorological uses of the technique, Bendat present some of the more practical aspects of the measurement of power spectra. The other references are concerned with the application of spectral techniques to the fields of meteorology, oceanography, and aeronautical engineering, and with more detailed discussions of the mathematical basis of spectral analysis. Although a knowledge of the theory underlying the technique of spectral analysis is certainly desirable, a lack of appreciation of the finer points of the mathematical development need not prevent the successful use of spectral analysis as a useful statistical tool in the solving of engineering problems.

The discussion here has been limited to the computation and interpretation of the spectrum of an individual time-series record. When two different records (perhaps a DO record and a BOD record) are analyzed together in spectral computation, the result, called the "cross-sprectrum," is a much more powerful tool than are the individual spectra. Cross-spectra, when combined with other spectral calculations, produce among other things quantitative information on the response of one record to another (the change of DO as a function of the change in BOD, for example) and on the time lag with which the response occurs. Although this type of information is certainly of considerable importance in engineering problems, this discussion has been directed toward presenting the foundation upon which the more esoteric spectral calculation rests.

ACKNOWLEDGMENTS

The author is deeply indebted to Dr. Blair Kinsman of the Chesapeake Bay Institute for supervising his initiation into the mysteries of spectral analysis.

The data discussed here were obtained by the Public Health Service during a field survey of the Potomac Estuary. This survey was conducted at the request of the U.S. Army Corps of Engineers as part of a comprehensive study of the water resources of the Potomac River Basin. The permission of the Corps of Engineers to use these data is gratefully acknowledged.

GPO 806-514-3

ENTS

...r Kinsman of the
...nitiation into the

... the Public Health
... Estuary. This sur-
... Army Corps of
... of the water re-
... mission of the
...fully acknowledged.

BIBLIOGRAPHY

1. Barber, N. F. Experimental Correlograms and Fourier Transforms. Pergamon Press, New York, N.Y., 1961.
2. Bendat, J. S. Principles and Applications of Random Noise Theory. Wiley, New York, N.Y., 1958.
3. Blackman, R. B. andJ. W. Tukey. The Measurement of Power Spectra. Dover, New York, N.Y., 1958.
4. Grenander, U. and M. Rosenblatt. Statistical Analysis of Stationary Time Series. Wiley, New York, N.Y., 1957.
5. Hannan, E. J. Time Series Analysis. Wiley, New York, N.Y., 1960.
6. Kinsman, B. Surface Waves at Short Fetches and Low Wind Speeds - A Field Study. Chesapeake Bay Inst., Johns Hopkins Univ., Tech. Rep. XIX, May 1960.
7. Marks, W. and W. J. Pierson. The power spectrum analysis of ocean-wave records. Trans. Am. Geophys. Union. 33:834-44. 1952.
8. Munk, W. H., F. E. Snodgrass, and M. J. Tucker. Spectra of low-frequency ocean waves. Bul. Scripps Inst. Oceanog. Univ. Calif. 7(4): 283-362. 1959.
9. Panofsky, H. A. Meteorological applications of power spectrum analysis. Bull. Am. Meteorol. Soc. 36:163-66. 1955.
10. Panofsky, H. A. and G. W. Brier. Some Applications of Statistics to Meteorology. Pennsylvania State University. 1958.
11. Press, H. and J. W. Tukey. Power Spectral Methods of Analysis and Their Application to Problems in Airplane Dynamics. Flight Test Manual, NATO, Advisory Group for Research and Development, IV-C, 1-41. June 1956.
12. Rice, S. O. Mathematical analysis of random noise. Bell System Tech. 23: 282-332. July 1944; 24:46-156. Jan. 1956. Reprinted in: Selected Paper on Noise and Stochastic Processes. N. Wax, ed. Dover, New York, N.Y., 1954.
13. Taylor, G. I. Statistical theory of turbulence. Proc. Roy. Soc. (London). A151:421-78. 1935.

GPO 806-514-2